DOCUMENT RESUME

ED 347 174 TM 018 540

AUTHOR Caldwell, Rebecca
TITLE Teaching the Effects of Additive and/or Multiplicative Constants on the Mean, Moments about the Mean, and Correlation.
PUB DATE Apr 92
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 20-24, 1992).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Mathematics; *Correlation; *Data Analysis; Mathematics Instruction; *Statistics; Teaching Methods
IDENTIFIERS *Additive and Multiplicative Constants; Additive Models; Kurtosis; *Mean (Statistics); Multiplier Effect; Skew Curves; Standard Deviation

ABSTRACT

Understanding the effects of additive and multiplicative constants on basic statistical procedures provides the important groundwork necessary for a thoughtful and logical understanding of more complex methods. This paper explores the effects of additive and multiplicative constants on the mean, standard deviation, skewness, kurtosis, and correlation. A constant is an operation consistently applied to all observations on a given variable. Small data sets and graphs are used to make the discussion more accessible and comprehensive. Additive and multiplicative constants have consistent and predictable effects on the mean, moments about the mean, and correlation. Three tables of data and seven graphs illustrate the analyses. Seven references are included. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
REBECCA L. CALDWELL
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED347174

Teaching the Effects of Additve and/or Multiplicative Constants on the Mean, Moments About the Mean, and Correlation

Rebecca Caldwell
Texas A&M University 77843-4225

Paper presented at the annual meeting of the American Educational Research Association, San Fransisco, CA, April 23, 1992.

TM018540

BEST COPY AVAILABLE

Abstract

Understanding the effects of additive and multiplicative constants on basic statistical procedures provides the important groundwork necessary for a thoughtful and logical understanding of more complex methods. This paper explores the effects of additive and multiplicative constants on the mean, standard deviation, skewness, kurtosis, and correlation. Small data sets and graphs are used to make the discussion more accessible and comprehensive.

Statistical equations are governed by mathematical logic that is easily overlooked once an individual is faced with complex formulas and computer printouts. In order to help demystify some of those equations and concepts, an exploration of the effects of mathematical operations, such as addition and multiplication, on basic statistical procedures, can be useful. All parametric methods, such as t-tests, ANOVA, ANCOVA, or regression, use weighting systems in the form of additive and multiplicative constants, so an understanding of their effects is important. Because all parametric methods can be treated as special cases of correlation analysis (Baggalay, 1981; Dolenz, 1992; Thompson, 1991), an understanding of the effects that additive and multiplicative constants have on the correlation coefficient will particularly enhance the learning of more advanced techniques. The effects of additive and multiplicative constants on the mean and moments about the mean (standard deviation, skewness, and kurtosis), and correlation, will be explored here through the manipulation of small data sets and graphs.

Constants

A constant, as implied by the name, is an operation consistently applied to *all* observations on a given variable. For example, in the regression equation (the equation of a straight line),

$$Y=a + b\,(x)$$

the additive and multiplicative constants, a and b respectively, are applied to *every* observation of variable X.

Standard scores are another example of the use of constants in statistics. Raw scores may be converted to a z scores using the formula:

$$X - \overline{X} \text{ / standard deviation.}$$

These scores are then often further manipulated by the use of additive and multiplicative constants to make the scores easier to use, as z scores entail the use of negative numbers and decimals. IQ scores are an example of z scores that have been converted in this manner, using the following formula:

$$100 + 15 \text{ or } 16 \text{ (z)},$$

in which the additive constant is 100, and the multiplicative constant is either 15 or 16. The mean of the converted score is 100, and the standard deviation is 15 or 16. The meaning of the converted score is no different from that of a z score with a mean of 0 and and standard deviation of 1; both are an equally good reflection of any given score's relation to other scores in terms of distance from the mean. But we may have stylistic preferences for scores to be expressed in certain ways.

Constants and the Mean

Constants have consistent and predictable effects on statistical equations. We can begin to explore this by looking at how an additive or multiplicative constant affects the mean. The mean is the arithmetic average of all scores on a given variable, and is represented by the formula,

$$\overline{X} = \Sigma Xs / n.$$

We can see by examining Table 1 that when the constant 2 is added to each score, the mean of that score increases by 2. The graphic

representation of this change is shown in Figure 1a and 1b. We can derive the following rule:

Adding a constant value to each score on a given variable results in a new mean which is equal to the original mean plus the constant value. Since addition and subtraction are really the same operation, the same holds true when a constant value is subtracted from each score; the new mean is equal to the old mean minus the value of the constant.

INSERT TABLE 1 ABOUT HERE

INSERT FIGURE 1 ABOUT HERE

If we multiply each of the original scores in our data set by a constant value, 2, as in Table 1, the new mean is equal to 6, or the old mean times 2. This change is shown graphically in Figure 1a and 1c. Thus, we can derive another rule:

When each score on a given variable is multiplied by a constant value, the new mean is equal to the old mean times that constant value. Since multiplication and division are really the same operation, the same holds true for division: the new mean equals the old mean divided by the constant value.

Constants and the Standard Deviation

There are several measures that reflect the way scores are spread out, including the sum of squares (SOS), variance, and standard deviation. The sum of squares is the sum of each score minus the mean, squared ($\Sigma (X - \bar{X})^2$), and reflects how much

deviation there is from the mean. The values of X - X are squared to eliminate negative numbers.

The variance is simply the sum of squares divided by n (for a population) or n-1 (for a sample), and is helpful because this calculation reduces the influence of group size. That is, when you have a larger n, your calculated SOS is likely to be larger, too. Calculating the variance removes the influence of n, because variance reflects average squared deviation from the mean per unit of X.

The standard deviation is the square root of the variance, which is useful because it converts the variance back to the original, nonsquared metric. If we have a variable measured in dollars, and the variance is 225 while the standard deviation is 15, the variance is 225 squared dollars, and the standard deviation is 15 dollars. We usually can more readily interpret statistics measured in the same metric of the variable, as with dollars against squared dollars. If we have a mean of 300 and a standard deviation of 15, then a raw score of 345 is 3 standard deviations from the mean, which is the 99.9th percentile. Here is a summary of the measures of spreadoutness we have just discussed:

$$SOS=\Sigma(X-\overline{X})^2$$

$$\text{Variance} = SOS/n \text{ (or } n-1)$$

$$\text{Standard deviation} = \sqrt{\text{variance}}$$

As we observed above, adding a constant to each score on a given variable does not change the way those scores are spread out, thus it does not change the standard deviation. However, multiplying each observation in a set of scores by a constant does

change the way the scores are spread out, and this change is reflected in the calculated standard deviation.

Multiplying by a negative constant does not change the sign of the standard deviation, because conceptually there cannot be less than zero spreadoutness. An examination of Table 1 in the case of the multiplicative constant shows that the new standard deviation is equal to our old standard deviation times two. We can deduce the following rules:

Adding or subtracting a constant to each of a set of scores does not change the standard deviation.

Multiplying each of a set of scores by a constant results in a new standard deviation, which is the old standard deviation times the absolute value of the constant.

Dividing each of a set of scores by a constant results in a new standard deviation, which is equal to the old standard deviation divided by the absolute value of the constant.

Negative multiplicative constants do not change the sign of the standard deviation.

The standard deviation is a reflection of how the scores on a given variable are spread about the mean, but when calculated using raw scores, it is also a reflection of the unit of measurement used. This is another way of saying that it is affected by multiplicative constants. For example, say that the variable we are measuring is height. We could choose to measure height in inches or centimeters; either would be appropriate. However, because inches are larger than centimeters (centimeter = .39 inch), the standard deviation for

the data as measured in inches will be larger than the data as measured in centimeters. This is shown in Table 2.
Because .39 is the multiplicative constant in this example, the standard deviation for centimeters is equal to .39 times the standard deviation for inches.

INSERT TABLE 2 ABOUT HERE

The Use of z Scores

As is the case with inches and centimeters, all studies in education and the behavioral sciences involve the use of measures, which at some level, are arbitrary. Z scores are a way of taking out the effect of the arbitrary measure by standardizing the standard deviation, which is affected by the multiplicative constant, as illustrated above. Again, the formula for converting a raw score to a z score is as follows:

$$z = X - \overline{X} / \text{standard deviation.}$$

Looking at Table 2, we can see that the z scores for the variable as measured in inches and centimeters are the same. Z scores always have a mean of 0 and a standard deviation of 1. Z scores are more informative than raw scores, because they include information about the mean and standard deviation (Kachigan, p. 64). If we have a z score of 2, we know that the mean is 0, and the score we are examining is exactly 2 standard deviations from the mean. Z scores can be positive or negative, depending upon which side of the mean they are on.

Constants and Skewness

The third and fourth moments about the mean, skewness and kurtosis, are calculated on the basis of scores that have been standardized, as in the case of the z scores above. Therefore, we could accurately expect that additive and multiplicative constants will have limited effects on either skewness or kurtosis. Skewness is a measure of the nonsymetry of a curve as it is viewed from left to right. In a skewed right, or positively skewed curve, there are comparatively more scores at the lower end of the scale of measurement than at the higher end. Conversely, in the skewed left, or negatively skewed curve, there are comparatively more scores clustered at the higher end of the scale of measurement than at the lower end (Hinkle, Wiersma, & Jurs, p. 40). Skewness can be calculated using the following formula:

$$\text{Skewness} = \sum z^3/n$$

A perfectly symmetrical, or unskewed curve, would have a skewness coefficient of 0. A curve with a skewness coefficient of 0 is never changed by additive or multiplicative constants. Neither is a skewed curve, except in the case of a negative multiplicative constant, which flips the curve to the other side of the 0 point when presented graphically, and changes the direction of the skew. One can visualize this by imagining that the Y-axis is the center around which the graph rotates 180°. This effect is shown in Figure 2.

INSERT FIGURE 2 ABOUT HERE

The absolute numerical value of the constant does not effect the skewness coefficient, it is only the sign of the constant that

changes it. This is illustrated in Table 2 by observing the change in skewness between the three conditions shown. In column 3 of Table 2, the data from column 2, centimeters, have been multiplied by -1 to show the effect of negative constants on the moments about the mean. (Obviously, this is for illustration, as there is no such thing as negative centimeters.) We can see that skewness for the data as measured in inches or centimeters is -.506, and the skewness for centimeters, after having been multiplied by -1, is +.506.

Further examination of the nature of skewness will make the reasons for this clearer. As a reminder, the normal curve is a kind of frequency distribution where the score on the variable we are interested in is on the X-axis, with the lower scores on the left, and the higher scores on the right. The Y-axis is a measure of frequency, or the number of times our subjects got a particular score. A skewed distribution means that there were more scores piled up on either the low or the high end instead of the middle for instance, as in a normal curve. Multiplying each score by a positive constant does not change the number of times each score is attained relative to the others, and thus does not change which side they are piled up on. However, if we apply a negative multiplicative constant to each score, the higher scores become the lower scores and the direction of the skew reverses.

The only exception to this pattern is if each score on our variable is multiplied by zero. Figuring skewness then would become an "illegal" mathematical operation because it would involve dividing by the standard deviation (when calculating z scores), which would be 0, and you cannot divide by 0. In this way, we know that

there must be some variability in scores in order to calculate skewness. From this we can make the following observations:

Adding or subtracting a constant from each score on a given variable does not change the coefficient of skewness for that variable.

Multiplying each score on a given variable by a constant value does not change the skewness coefficient when the curve is not skewed (the skewness coefficient is 0). When there is some skew (the skewness coefficient is not equal to 0), a multiplicative constant will not change the skewness coefficient unless the constant is a negative number. When the multiplicative constant is negative, the absolute numerical value of the skewness coefficient does not change, only the sign of the skewness coefficient changes. In the case that the multiplicative constant is zero, skewness cannot be calculated because it is an "illegal" mathematical operation.

Constants and Kurtosis

Kurtosis, the fourth moment about the mean, is a measure of the tallness or flatness of a curve. The kurtosis of a curve is measured as it differs from the height of the normal curve. A tall curve is called leptokurtic, a normal one is mesokurtic, and a flat curve is platykurtic (Runyon & Haber, 1984). Kurtosis is calculated using the following formula:

$$\text{Kurtosis} = (\Sigma z^4/n) - 3$$

The kurtosis coefficient for a normal curve is 0. A leptokurtic curve has a kurtosis coefficient greater than zero, and a platykurtic curve

has a kurtosis coefficient less than 0 (Bump, 1991). Kurtosis is not affected by either additive or multiplicative constants. Again, we can see why this is the case by examining the nature of kurtosis.

Kurtosis is a facet of a frequency distribution, or the way our scores pile up in relation to each other. When we have a tall curve it means that more of our subjects got a score in the midrange, and fewer at the high and low ends than is the case in a normally shaped curve. Adding or multiplying each score by a constant does not change the relative frequencies of each score. Even if we have a negative multiplicative constant, the height of the curve is not changed by the advent of the opposite-signed score, as we can see in Table 2 in comparing the kurtosis coefficients for the three conditions. The only exception to this is if the multiplicative constant is zero. As is the case with skewness, figuring kurtosis would involve dividing by zero, which is an illegal mathematical operation. The following rules follow:

When a constant is subtracted or added to each score on a given variable, it does not change the kurtosis coefficient for that variable.

When each score on a given variable is multiplied by a constant, it does not change the kurtosis coefficient for that variable.

When the multiplicative constant is zero, calculating kurtosis becomes an "illegal" mathematical operation.

Constants and Correlation

Correlation (r) is the degree of relationship between two variables. This means that when scores on one variable go up, scores on the other variable go up too, or when scores on one variable go down, the scores on the other variable go down, also. In the case of a negative correlation, when scores on one variable go up, the scores on the other variable go down. In each case the scores are related to each other in a systematic way. The formula for correlation is as follows:

$$r_{xy} = \Sigma z_x z_y \ /n-1$$

We can see by looking at Table 3 that when the additive and multiplicative constants were applied to our variable B they did not change the relationship, or correlation, between the variables. However, where the multiplicative constant was negative, it changed the direction of the relationship.

INSERT TABLE 3 ABOUT HERE

Figure 3a is the graph of a perfect correlation. We can see that for each *high* X score, there is a correspondingly *high* Y score. When each Y score is multiplied by a negative constant, the appearance of the graph changes to Figure 3b, where for each *low* X score, there is a correspondingly *high* Y score. The extent of the relationship, or the absolute value of r, does not change, it is only the direction of the relationship that changes. If each score on one or both variables is multiplied by the constant 0, however, the operation of calculating r becomes illegal because it involves dividing by 0. This belies the

assumption that in order to calculate correlation, there *must* be two true *variables*. We can derive the following rules:

Subtracting or adding a constant to each score on a given variable does not affect correlation.

Multiplying each score on a given variable by a constant value does not change the correlation coefficient unless the constant is a negative number. When the multiplicative constant is negative, the absolute numerical value of the correlation coefficient does not change, only the sign of the correlation coefficient changes.

In the case that the multiplicative constant is zero, correlation cannot be calculated because it is an "illegal" mathematical operation.

INSERT FIGURE 3 ABOUT HERE

Summary

Additive and multiplicative constants have consistent and predicable effects on the mean, moments about the mean, and correlation. Understanding constants is an integral part of a true understanding of statistics, as their effects are part of the reason why some procedures, such as calculating standardized scores, have been derived. This paper has explained the nature of constants by the use of discussion, small data sets, and graphs.

References

Baggalay, A.R. (1981). Multivariate analysis: An introduction for consumers of behavioral research. Evaluation Review, 5, 123-131.

Bump, W. (1991, January). The normal curve takes many forms: A review of skewness and kurtosis. Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, TX.

Dolenz, B. (1992, January). Factors that attenuate the correlation coefficient and its analogs. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, TX.

Hinkle, D.E., Wiersma, W., & Jurs, S.G. (1988). Applied statistics for the behavioral sciences. Boston: Houghton Mifflin Company.

Kachigan, S.K. (1986). Statistical analysis: An interdisciplinary introduction to univariate and multivariate methods. New York, Radius Press.

Runyon, R.P., & Haber, A. (1984). Fundamentals of behavioral statistics. New York: Random House.

Thompson, B. (1991). A primer on the logic and use of canonical correlation analysis. Measurement and Evalation in Counseling and Development, 24 (2), 80-95.

Table 1

	Original	Plus 2	Times 2
	1	3	2
	2	4	4
	3	5	6
	4	6	8
	5	7	10
Mean	3	5	6
Standard Deviation	1.58113883	1.58113883	3.16227766

Table 2

	Inches	Centimeters (Inches (.39))	Centimeters (-1)
	8	3.12	-3.12
	8	3.12	-3.12
	7	2.73	-2.73
	8	3.12	-3.12
	12	4.68	-4.68
	11	4.29	-4.29
	10	3.9	-3.9
	9	3.51	-3.51
	7	2.73	-2.73
	4	1.56	-1.56
	2	0.78	-0.78
	1	0.39	-0.39
Mean	7.25	2.8275	-2.8275
Standard Deviation	3.387812381	1.321246828	1.321246828

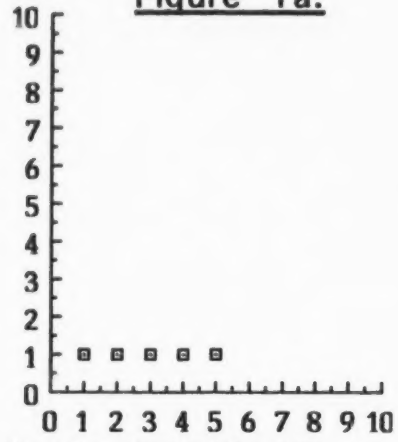
	Z scores: inches	Z scores: centimeters	Z scores: centimeters (-1)
	0.221381799	0.221381799	-0.221381799
	0.221381799	0.221381799	-0.221381799
	-0.073793933	-0.073793933	0.073793933
	0.221381799	0.221381799	-0.221381799
	1.402084728	1.402084728	-1.402084728
	1.106908996	1.106908996	-1.106908996
	0.811733264	0.811733264	-0.811733264
	0.516557531	0.516557531	-0.516557531
	-0.073793933	-0.073793933	0.073793933
	-0.95932113	-0.95932113	0.95932113
	-1.549672594	-1.549672594	1.549672594
	-1.844848326	-1.844848326	1.844848326
Mean	0	0	0
Standard deviation	1	1	1
Skewness	-0.505524973	-0.505524973	0.505524973
Kurtosis	-0.993662794	-0.993662794	-0.993662794

Table 3

A	B	C	D	E
1	1	3	2	-2
2	2	4	4	-4
3	3	5	6	-6
4	4	6	8	-8
5	5	7	10	-10

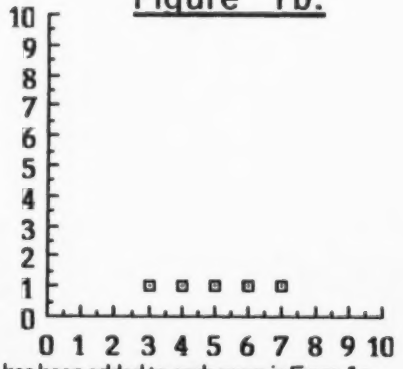
r of AB =1
r of AC =1
r of AD =1
r of AE = -1

Figure 1a.



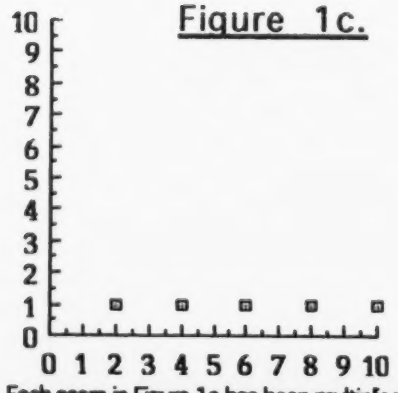
Mean=3, standard deviation=1.58.

Figure 1b.



2 has been added to each score in Figure 1a.
Mean=5, standard deviation=1.58.

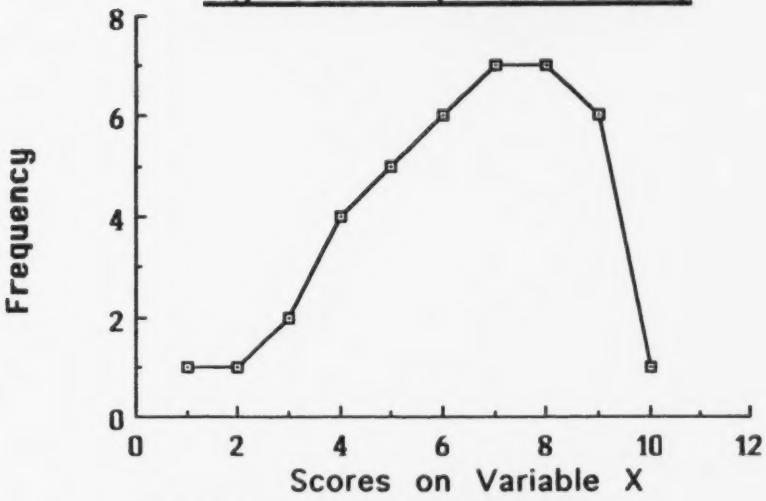
Figure 1c.



Each score in Figure 1a has been multiplied by 2.
Mean=6, standard deviation=3.16.

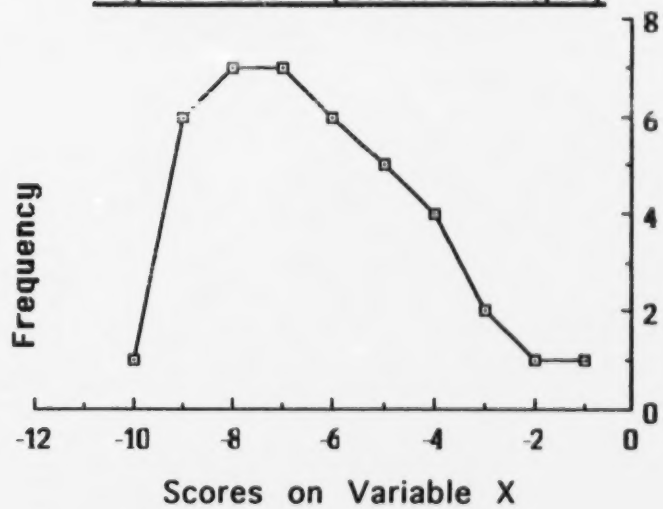
Adapted from Kachigan, 1986, p. 66.

Figure 2a. (Skewed Left)



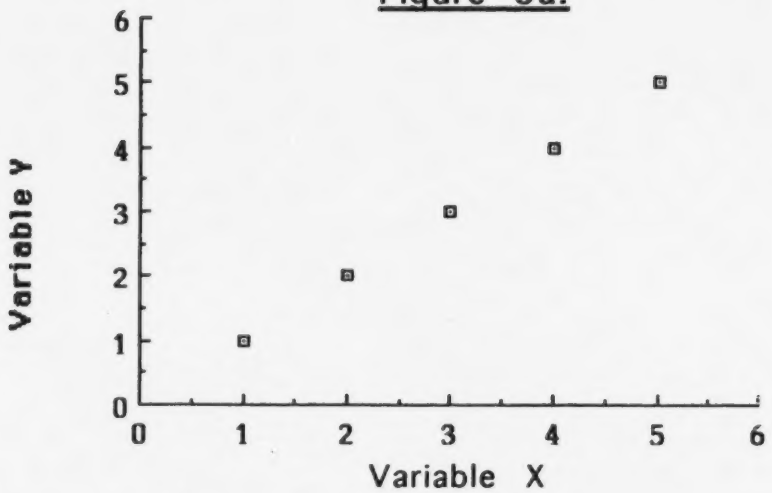
Data chosen by author for purposes of illustration.
Skewness coefficient= -.49.

Figure 2b. (Skewed Right)



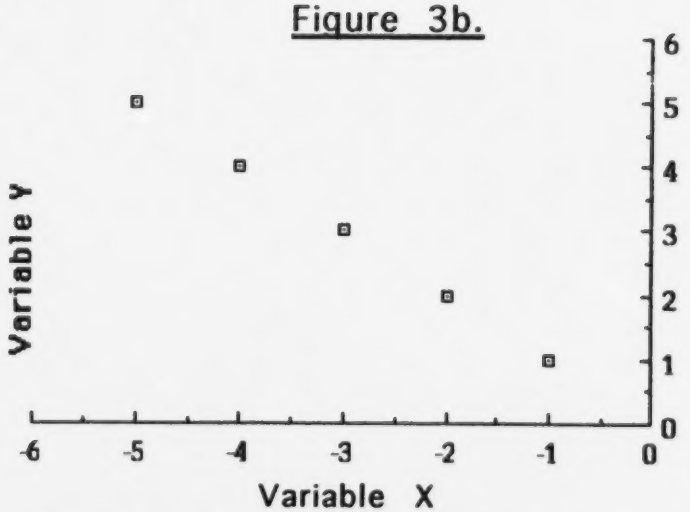
Scores in 2a have been multiplied by -1.
Skewness coefficient=.49.

Figure 3a.



Correlation (r)=1.

Figure 3b.



Scores on variable X in Figure 3a have been multiplied by -1.
Correlation (r)=-1.